The Impact of Trends in Technology on File System Design

Michael D. Dahlin
University of California, Berkeley
dahlin@cs.berkeley.edu

Abstract

This paper describes several key trends in technology that will influence the design of file systems for the next decade. It first outlines the basic trends to hardware performance that underlie file system design. These trends affect both user demands on file systems and trade-offs in their design. It then considers how technologies will drive more demanding file system workloads that will require scalable file systems. Next, it outlines how opportunities raised by these low-level technology trends impact specific aspects of the xFS file system's serverless design [Dahlin et al., 1994b, Anderson et al., 1996]. Finally, to put xFS's serverless approach in context, it discusses other technology trends that affect file systems but that are not explicitly addressed in the serverless design.

1. Trends in Technology

Table 1 summarizes the current performance and expected rates of improvement for hardware technologies that influence file system design. The data in this table provide a basis for four gen-

Hardware Parameter	1995 Baseline	Yearly Improvement Rate
Disk Latency	12 ms	10%/year
Disk Bandwidth	6-9 MB/s	20%/year
Disk Cost/Capacity	$0.24/MB	100%/year
Network Latency	1 ms	20%/year
Network Bandwidth	20 MB/s	45%/year
Processor Performance	100 SPECInt92	55%/year
Memory Bandwidth	30-70 MB/s	40%/year
Memory Cost/Capacity	$31/MB	45%/year

TABLE 1.**Summary of hardware improvement trends.** The 1995 Baseline reflects the performance a high-end, desktop workstation in 1995, and the Yearly Improvement Rate's reflects trends discussed in greater detail in Section 1. Although network latency has historically improved slowly as indicated in the table, research in low latency network protocols may provide dramatic improvements in the future. Conversely, while memory cost has historically improved quickly, recent progress has been much slower. This paper documents the rates of improvement.

eral sets of observations. First, disks are slow, and their performance improves slowly relative to other technologies. Second, compared with disks, networks have somewhat better bandwidth and much better latency, and both latency and bandwidth are improving more quickly for networks than for disks. Third, processors and memories are much faster than either disks or networks, and their performance is improving quickly. Fourth, disk storage is two orders of magnitude less expensive than equivalent memory storage, and this gap is widening as disk costs improve more rapidly than memory costs. The rest of this section describes these trends in detail.

1.1. Disk Performance

Disk accesses rely on mechanical motions that are slow compared to electronic operations such as DRAM reads. Further, the rate of improvement in disk mechanical performance is lower than the rate of improvement of integrated circuit performance. Disk delays are dominated by three factors: seek time, rotational latency, and transfer time. During the seek time, an actuator moves the disk heads to the disk cylinder being accessed. Rotational latency allows a specific disk block to spin under the disk head, and transfer time allows data to be accessed as it passes under the disk head. Note that for physically sequential transfers, only the first block need pay seek and rotation times.

Table 2 summarizes disk performance trends for high-performance, commodity disks appropriate for workstation servers in 1987, 1990, and 1994. The 1987 and 1990 values reflect performance parameters reported by Gibson [Gibson, 1992] while the information for the more recent drive is from a product data sheet [Seagate, 1994].

During this period, seek and rotational latencies have improved slowly, at about 10% per year, while bandwidth has increased more quickly, at about 20% per year. These improvement rates are similar to those discussed by Hennessy and Patterson [Hennessy and Patterson, 1996]. Bandwidth increases more quickly than rotational speed because of increasing storage densities on the disk — more data spins under the disk head at a given rotational speed.

A consequence of disks' high seek time and rotational latency is that large transfers use disks much more efficiently than small ones. As the table indicates, latency dominates the time to access 8 KB from disk while disk bandwidth dominates the time to transfer 1 MB. However, because disk bandwidths improve more rapidly than other aspects of disk performance, increasingly large transfers are needed over time to maintain a given level of efficiency. For instance, to reduce seek and rotational latency to less than 10% of the total access time required a 562 KB sequential transfer in 1987 and a 988 KB transfer in 1994.

1.2. Disk Capacity

Disk capacity per unit cost improves dramatically over time — at 100% per year since 1992. As the data in Figure 1 indicate, between 1983 and 1995 the cost of disk storage has fallen by more than a factor of 400, from about $100/MB in January of 1983 to less than $0.24/MB in July of 1995. For instance, advertisements in *Byte* magazine in January of 1983 offer a 44 MB hard drive for $4,395 while advertisements in the July 1995 edition price a 9,000 MB drive at $2,159. This improvement reflects an increase of 62% per year over the twelve and one-half year period with a faster rate since 1992 as PC manufacturing volumes and competition have ramped up.

1.3. Network Latency

Between 1989 and 1994 network RPC latency improved by less than a factor of three, from 2.66 ms [Schroeder and Burrows, 1990] to about 1.00 ms, a speedup of about 20% per year. However, a number of research efforts have demonstrated extremely low network latency using proto-

				Improvement Rates		
	1987	**1990**	**1994**	**1987-94**	**1987-90**	**1990-94**
Model	Fujitsu M2361A	Seagate ST-41600n	Seagate ST-15150n			
Average Seek Time	16.7 ms	11.5 ms	8.0/9.0 ms (rd/wr)	11%/yr	13%/yr	9%/yr
Rotational Speed/Avg. Latency	3600 rpm/8.3 ms	5400 rpm/5.5 ms	7200 rpm/4.2 ms	10%/yr	15%/yr	7%/yr
Bandwidth	2.5 MB/s	3-4.4 MB/s	6-9 MB/s	20%/yr	21%/yr	20%/yr
8 KB Transfer	28.3 ms	18.9 ms	13.1 ms	12%/yr	14%/yr	10%/yr
1 MB Transfer	425 ms	244 ms	123 ms	19%/yr	20%/yr	19%/yr

TABLE 2.**Disk performance parameters.** This table compares the parameters of a 1987, 1990, and 1994 commodity disk drive. Although these disks represent only a single data point for each year, I believe the trends accurately describe high-end commodity hardware available over the specified periods. Average seek time is the time it takes to seek one-third of the way across the disk surface, so it represents a pessimistic estimate of the seek time for a typical request because real requests will benefit from locality [Hennessy and Patterson, 1996]. For the 1994 disk, the seek time is higher for writes than for reads because of more stringent head settling requirements for writes. The rotational latency represents the time for the disk to complete half of a complete revolution at the given rotation speed. The bandwidth is the uncached transfer rate for a large data transfer, ignoring seek and rotational latencies. For the 1990 and 1994 drives, transfer bandwidth is higher near the edges of the disk than near the center because of the higher linear velocity of the media near the edges. To provide an upper bound on disk improvement rates, I use the more aggressive read seek time and outer-edge bandwidths when computing improvements over time. The 8 KB and 1 MB transfer times indicate the time to transfer an 8 KB or 1 MB contiguous block of data including the time for one "average" seek and rotational latency, plus the transfer size divided by the disk's maximum bandwidth.

cols somewhat more restrictive than those of RPC. Thekkath, for example, demonstrated a restricted RPC protocol with 72 µs round-trip latency on 155 Mbit/s ATM [Thekkath and Levy, 1993], and Martin and von Eicken have measured 29 µs and 52 µs round trip times for active message implementations using FDDI and ATM, respectively [Martin, 1994, von Eicken et al., 1995]. If any of these simplified protocols could be used as a fast path for file system communication, order of magnitude improvements in network latencies may be possible.

Whether or not research efforts provide new protocols, network latencies are significantly smaller than disk latencies, and they seem likely to improve more quickly in the future. As a result, networks will provide faster access to small data items than will disk.

1.4. Network Bandwidth

Over the past 15 years, the aggregate network bandwidth available to a 32-node cluster of high-end, desktop workstations has increased by a factor of 256, from a 10 Mbit/s shared Ethernet in the early 1980's to a 2,560 Mbit/s switched, Myrinet [Boden et al., 1995] network in which half

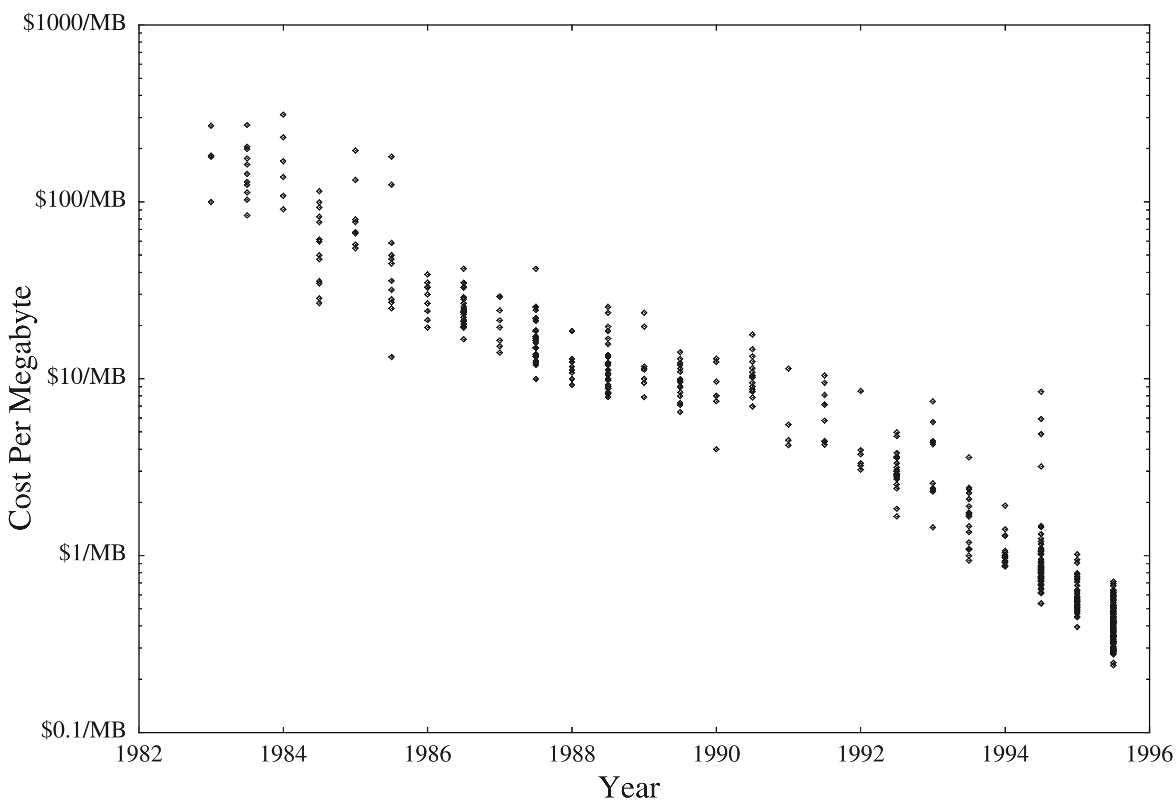

FIGURE 1.**Disk cost expressed in dollars per megabyte of capacity.** Each point indicates the capacity in megabytes divided by price in dollars for a hard disk drive advertised in the January or July edition of *Byte* magazine for the indicated year. Note the log scale for the y-axis.

of the workstations can each send at 160 Mbit/s while half receive at the same rate. This improvement represents an increase of about 45% per year over this period. This improvement in network speed has come from two sources, faster network links to each machine, and better topologies to increase the aggregate bandwidth above the per-link bandwidth. The per-link bandwidth has increased by a factor of 16 over this period, at about 20% per year. The other factor of 16 comes from the move from the bus-based Ethernet to switch-based topologies [Schroeder et al., 1991] such as ATM [ATM Forum, 1993], Myrinet, and switched Ethernet.

Although the one-time move from bus-based topologies to switch-based topologies contributed significantly to the improvement in network bandwidths over the last decade, large improvements in network speed are likely to continue because it is easier to increase the speed of switched network links than bus-based network links. First, many switched networks allow individual links to be upgraded independently; in the past, a network's speed could only be increased if all of the machines in a system were simultaneously upgraded. Second, electrical signalling is easier on these point-to-point networks; the bandwidth of the Myrinet network, for instance, is limited by the workstation's I/O bus bandwidth rather than the 640 Mbit/s physical link bandwidth [Martin, 1995].

This evaluation suggests that high-performance networks will be faster than high-performance disks, even for large data transfers.

1.5. Processor Performance

Processor performance benefits from improving integrated circuit technologies that allow designers to increase the number of transistors per chip and to increase the speed of these transistors. As a result, processor speed increases at 50% to 55% per year [Gwennap, 1995].

1.6. Memory Bandwidth

Improvements to DRAM architectures — including wider memories, memory interleaving, and efficient access modes [Bursky, 1992, Jones, 1992, Prince et al., 1992, Hart, 1992, Bondurant, 1992] — allow memory bandwidths to improve quickly. Table 3 summarizes the memory-to-memory copy bandwidths of machines available in 1989 compared to machines measured in late 1994 and suggests that memory bandwidths are increasing at 40% per year or better for each product line. In absolute terms, memory speeds are much faster than disk speeds, and the relatively high rate of memory speed improvements suggest this gap will remain large.

1.7. Memory Capacity

The capacity of memory chips increases at about 60% per year due to improvements in technology that quadruple the amount of memory that can fit on a chip every three years [Hennessy and Patterson, 1996]. As Figure 2 indicates, before 1992, memory prices reflected this trend, falling at over 100% per year during that time. During the last three years, however, the cost per megabyte of memory has been flat. It is not yet clear whether this drastically reduced rate of improvement in cost is a long-term effect or a temporary blip. Over the full seven-year period indicated in the figure, prices fell by an average of about 45% per year, but the most recent trends suggest that this may be an optimistic estimate for memory-size improvement rates. In any event, memory capacity seems likely to grow more slowly than disk capacity in the future.

2. Increasing Demands on File Systems

A key trend that impacts file systems is the dramatically increasing computing power available to users. These increases come from several sources. New, more powerful CPUs increase the demands of sequential applications, and fast networks give users the ability to exploit multiple processors to run distributed or parallel applications, further increasing both average and peak demands on file systems. Furthermore, expanding computing power enables new classes of demanding applications such as databases and multimedia programs that add additional pressure to I/O systems. As a result, I/O systems must advance quickly to avoid limiting system performance. This trend motivates the radically more scalable serverless design.

	1989		1994		
Manufacturer	**Machine**	**Bcopy Bandwidth**	**Machine**	**Bcopy Bandwidth**	**Five-year Average Rate of Increase**
SUN	SPARCStation 1	5.6 MB/s	SPARCServer 20 Model 6	27.9 MB/s	38%/year
Hewlett-Packard	9000-835CHX	6.2 MB/s	715/80	32.1 MB/s	39%/year
IBM	RT-APC	5.9 MB/s	6000/370	67.2 MB/s	63%/year
Digital Equipment	DS3100	5.1 MB/s	3000/400	43.8 MB/s	54%/year

TABLE 3.**Memory bandwidth for 1989 and 1994 machines.** The 1989 values were measured by Ousterhout [Ousterhout, 1990]. The 1994 values are for large, 10 MB transfers to reduce cache effects. The rate of increase is the change for each manufacturer for the products listed computed over five years. Note that these rates would differ somewhat if different products in each vendor's line were chosen.

2.1. Demands From Sequential Applications

The faster improvement rate of processor performance compared to that of disk performance suggests that future I/O systems be designed to take advantage of multiple-disk transfers; improving computers allow users to process more data more quickly, increasing the speed required from the file system. For example, the Case/Amdahl rule of thumb suggests that a balanced machine should have 1 Mbit of I/O bandwidth for each MIPS of CPU performance [Hennessy and Patterson, 1996]. Figure 3 plots the actual local single-disk bandwidth against processor performance for ten machines manufactured during the last decade. This graph suggests that high-performance machines with single disks no longer conform to the Case/Amdahl rule of thumb and that the gap is growing larger as processors get faster. The trend discussed above, whereby processors get faster more quickly than disk transfers, suggests that this imbalance will continue to grow.

If anything, Figure 3 understates the imbalance. First, it assumes that disk performance improves with disk bandwidth. In fact, because disk latency improves more slowly than disk bandwidth, overall disk performance is likely to improve more slowly as well. Second, as relative disk latencies increase, systems use more aggressive prefetching [Patterson et al., 1995], which

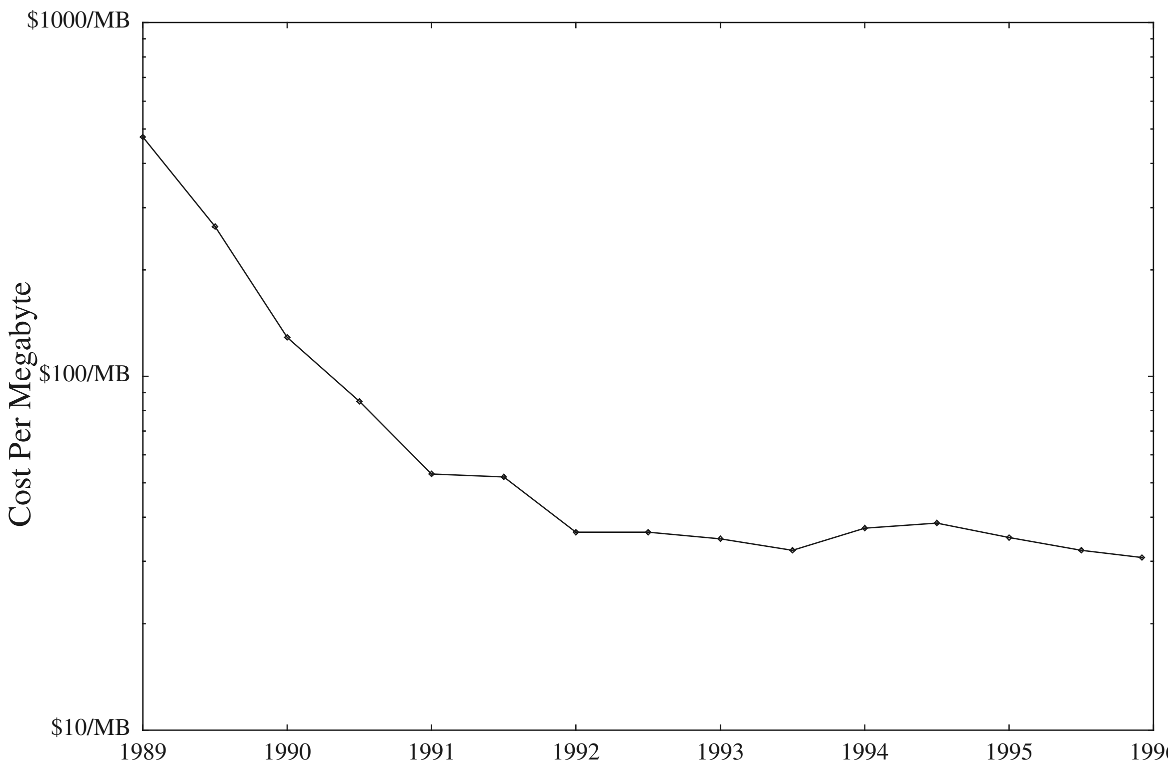

FIGURE 2. **DRAM price per megabyte.** Each point indicates the cost per megabyte of DRAM memory based on advertisements from *Byte* magazine. All prices reflect memory purchased in 9-bit-wide SIMM form with access times of 80 ns or better. To provide the best price for each data point, points before July of 1991 are the prices for 1MB SIMMs, and points from July of 1991 and later are one quarter of the prices of 4 MB SIMMs. Note the log scale for the y axis.

increases demands for disk throughput. In effect, prefetching reduces latencies for disk requests by increasing demands for bandwidth.

2.2. Demands From Parallel and Distributed Applications

Although sequential processors stretch the limits of conventional file systems, the increasing need for high-performance, parallel file access by mainstream users seems likely to swamp them. Efforts to allow users to harness the aggregate power of many workstations in Networks of Workstations (NOWs) [Anderson et al., 1995] will increase both the average and peak demands on file systems from parallel programs or multiple, simultaneous, sequential jobs.

The ability to harness multiple machines will allow users to tackle larger problems, increasing the total amount of data passed through file systems. Cypher [Cypher et al., 1993] measured the I/O demands of parallel supercomputing applications and found that average I/O requirements increased linearly as the size of a problem increased. This conclusion suggests that the perfor-

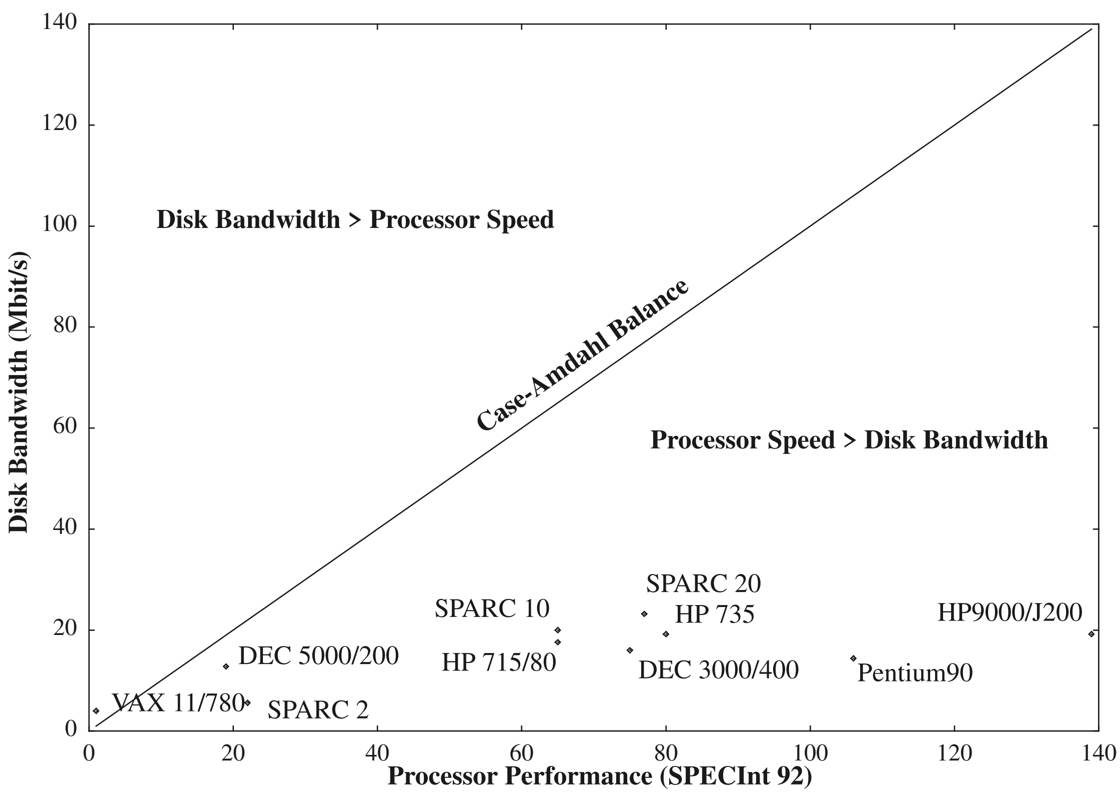

FIGURE 3. **Bandwidth and processor performance.** This comparison of local disk bandwidth against processor performance suggests that single-disk sequential machines have too little I/O capacity for their processing power and that this imbalance is increasing as processors get faster. Each point shows the measured bandwidth when reading a file that is too large to fit in the in-memory cache from a single local disk plotted against the published SPECInt performance for the machine. Note that SPECInt92 is used to rate processor power rather than MIPS. If the higher MIPS figures had been used, the machines would appear to be even more out of balance.

mance I/O systems must scale as quickly as that of aggregate — not just sequential — processing for systems that support parallel jobs.

Further, not only will NOWs increase total demands on file systems, these demands are likely to come in bursts as the actions of many machines become correlated. Where traditional file systems could depend on the "law of large numbers" to smooth file system demands over time, parallel and distributed jobs running on NOWs synchronize the actions of many machines. For instance, NOW users may compile programs using parallel `make` rather than sequentially. Even if a parallel `make` demands no more total work from the file system that a sequential one does, requests are sent to the file system much more quickly in the parallel case. To keep from reducing perceived end-user performance, the file system must provide higher peak performance to keep up with increased peak demands from users.

2.3. Demands From New Applications

Increasing workstation capabilities also enable new classes of demanding applications. For instance, I/O intensive database applications would challenge conventional file systems because they require large numbers of relatively small, concurrent accesses. Further, database vendors want access to a large number of disks and they want to control how data are distributed over those disks to assist in load balancing.

Continuous media applications such as video and audio will also pressure file systems by requiring them to supply large amounts of aggregate bandwidth. Video applications demand significant bandwidths for periods ranging from seconds to hours, requiring that storage systems accommodate large numbers of simultaneous, demanding users. Furthermore, the sequential nature of video streams may prevent file system caching from being useful, although by the same token prefetching may be effective. Finally, continuous media applications may require real-time guarantees for smooth playback. Although this thesis does not address continuous media applications or real-time guarantees explicitly, it does suggest methods by which storage systems may be built to provide large amounts of aggregate bandwidth. These techniques could then be used as a basis for a continuous media file system.

3. Implications For File System Design

Not only do changing technologies increase the demands on file systems, they also change the design trade-offs that should be made to satisfy those demands. Fast, switched, local area networks such as ATM and Myrinet enable serverless file systems by allowing LANs to be used as I/O backplanes, harnessing physically distributed processors, memory, and disks into single file systems. The switched networks provide aggregate bandwidths that scale with the number of machines on the network, and new, low latency network interfaces enable closer cooperation among machines than has been possible in the past. As the rest of this section describes, the trend towards faster, scalable networks motivates each of the pieces of the serverless design, although particular design decision were also shaped by other technology trends.

3.1. Cooperative Caching

The serverless design replaces central server caching with cooperative caching that coordinates the clients' caches to allow reads not satisfied by one client's cache to be supplied from another's. Trends in network performance and in memory capacity motivate cooperative caching [Dahlin et al., 1994b].

High-speed, low-latency networks provide the primary motivation for cooperative caching by allowing clients to access remote memory much more quickly than they can access remote or even local disk, as Table 4 indicates. Where fetching data from remote memory might be only three times faster than getting the data from remote disk on an Ethernet, remote memory may now be accessed ten to twenty times more quickly than disk, increasing the payoff for cooperative caching. At the same time, fast networks reduce the effectiveness of architectures like AFS [Howard et al., 1988] that use local disk as a cache rather than using remote memory; with fast networks, such an approach will be much slower than cooperative caching.

The recent lag in improvements to memory capacity provides a second motivation for cooperative caching. As memory capacity grows more slowly, designers can no longer rely on growing memories to maintain or improve hit rates or to cope with the larger working set sizes enabled by larger disks. Instead, designers must use available memory more efficiently. Cooperative caching does so by providing improved global hit rates with a given amount of memory and by reducing duplicate caching between clients and the server and among clients.

3.2. Distributed Storage

A serverless system replaces centralized disk storage with storage servers that distribute storage to different machines' disks. This approach works well because of high-bandwidth networks. Furthermore, to exploit trends in disk technologies the serverless storage system redundant, log-structured storage [Hartman and Ousterhout, 1995] to provide availability and improve write performance [Anderson et al., 1996].

Fast, scalable networks motivate distributed disk storage. When networks are faster than disks, a single client can read or write at its full network bandwidth by accessing multiple disks that are distributed on the network. Additionally, because a switched network's aggregate bandwidth can be orders of magnitude larger than its per-link bandwidth, the peak aggregate bandwidth of a distributed disk system is the sum of all of the disk bandwidths or network link bandwidths in the system. In contrast, a centralized system's bandwidth is limited to the network link bandwidth of a single machine — the central server — even if the server uses a local RAID to increase its disk bandwidth.

The serverless storage system design stores redundant data on disks to provide high availability. The low cost of disk storage makes this approach to availability more cost-effective than alternatives designs such as replicated servers [Walker et al., 1983, Kazar, 1989, Popek et al., 1990, Liskov et al., 1991, Kistler and Satyanarayanan, 1992, Birrell et al., 1993].

			Ethernet		**77 Mbit/s ATM (Half of 155 Mbit/s)**	
	Local Memory	**Local Disk**	**Remote Memory**	**Remote Disk**	**Remote Memory**	**Remote Disk**
Mem. Copy	250 us	250 us	250 µs	250 µs	250 µs	250 µs
Net Overhead	--	--	400 µs	400 µs	400 µs	400 µs
Data	--	--	6250 µs	6250 µs	800 µs	800 µs
Disk	--	14,800 us	--	14,800 µs	--	14,800 µs
Total	**250 us**	**14,800 us**	**6,900** µs	**21,700** µs	**1450** µs	**16,250** µs

TABLE 4.**Cache miss time.** Time to service a file system's local cache miss from remote memory or disk for a slow network, Ethernet, and a faster network, a 155 MBit/s ATM network that achieves half of its maximum link throughput. Copy time for local memory is the measured time it takes to read 8 KB from the file cache on a DEC AXP 3000/400. Network overhead times indicate round-trip, small-packet latencies based on TCP times reported in [Martin, 1994] for a Hewlett-Packard 9000/735 workstation. Transfer figures for Ethernet make the unrealistically optimistic assumption that data is transferred at the full 10 Mbit/s link speed (in reality, because Ethernet utilizes a shared-bus architecture, transfer times could often be several times those listed above). The ATM transfer time assumes that half of the full 155 Mbit/s bandwidth is attained to account for protocol overheads [Keeton et al., 1995]. The disk transfer time is based on measured, physical-disk time (excluding queueing) for the fastest of three systems measured under real workloads by Ruemmler and Wilkes [Ruemmler and Wilkes, 1993].

Trends in disk technology also favor the log-structured storage design used in the storage servers. Log structured file systems [Rosenblum and Ousterhout, 1992, Seltzer et al., 1993] commit modifications to disk using large, contiguous writes to the log; this approach uses disks efficiently by amortizing seek and rotational delays over large writes. Furthermore, log structured file systems are more efficient when free disk space is plentiful [Rosenblum and Ousterhout, 1992]; cheap disks make it feasible to provide extra disk space to get improved performance.

3.3. Distributed Management

A serverless file system implements data management separately from data storage. Although this separation can add an additional network hop to access data compared to a central server, low-latency, scalable networks mitigate this cost. Low-latency networks make the additional network hops needed to locate data inexpensive compared to the network or disk transfer of the data itself, and scalable networks prevent congestion that could interfere with low-latency communication.

4. Other Technology Trends

A number of other technologies will also influence file systems during the next decade. The increasing distribution of data over wide area networks (WANs), the growing use of portable computers, and the emergence of robotic, tertiary storage libraries will all demand new file system capabilities. This thesis focuses on file service within the context of a LAN with on-disk storage and does not address the opportunities and requirements of these other technologies in detail. To put this thesis in context, this section summarizes the relationship of the ideas it examines to these wider technology trends.

The World Wide Web (WWW) enables data to be distributed widely and shared, and I believe that there are many interesting research questions related to using file system techniques to improve the efficiency and convenience of the WWW. However, because WAN and LAN performance, availability, security, and cost characteristics are so different, the appropriate trade-offs for WAN and LAN file system protocols will be different. As a result, I believe that future WAN file systems will implement two sets of protocols, one for communication within a cluster of machines on a fast LAN network and the second for communication among these clusters over a slower WAN [Sandhu and Zhou, 1992]. Although the research for this thesis is devoted to file systems within a LAN, the techniques described here would also be appropriate for the LAN-specific protocol in a LAN/WAN file system. I have examined protocols appropriate for WAN file systems elsewhere [Dahlin et al., 1994a].

This thesis also focuses on the needs of desktop, as opposed to portable, computers. Portable computers pose a number of challenges to file systems stemming from their frequent disconnection from the network [Kistler and Satyanarayanan, 1992] or their use of relatively low-bandwidth, wireless network connections [Le et al., 1995, Mummert et al., 1995]. Because the techniques described in this thesis assume an environment where machines are tightly coupled using a fast LAN, they would not work well as the primary file system for portable computers. In a mixed environment of portable and sedentary computers, however, I would envision these techniques being used on stationary workstations to provide a scalable file system "backbone." In such an environment, the portable machines would run a different protocol such as Coda [Kistler and Satyanarayanan, 1992] or Ficus [Popek et al., 1990], using this scalable backbone as their "server."

Finally, tertiary-storage, robotic libraries [Drapeau, 1993] provide an opportunity for large amounts of data to be stored more cheaply than they can be stored on disk. This advantage exists because these libraries allow massive archives to be built and some system services, such as file backup and restore, to be automated. Because the latency that results when these devices are accessed can be tens or hundreds of seconds, file systems must be carefully structured to hide this latency. Many of the techniques in this thesis address issues related to moving data among different storage servers and might be extended as a mechanism to provide migration to tertiary storage libraries. This thesis does not, however, examine the critical issue of migration *policy* — when to move data between storage levels to mask latency.

Appendix: Raw Data

The data plotted in Figure 1 (disk cost and capacity over time) and Figure 2 (DRAM cost and capacity over time) are available at http://www.cs.berkeley.edu/~dahlin/techTrends.

Bibliography

Anderson, T., Culler, D., Patterson, D., and the NOW team (1995). A Case for NOW (Networks of Workstations). *IEEE Micro*, pages 54–64.

Anderson, T., Dahlin, M., Neefe, J., Patterson, D., Roselli, D., and Wang, R. (1996). Serverless Network File Systems. *ACM Transactions on Computer Systems*, 14(1).

ATM Forum (1993). *The ATM Forum User-Network Interface Specification, version 3.0*. Prentice Hall Intl., New Jersey.

Birrell, A., Hisgen, A., Jerian, C., Mann, T., and Swart, G. (1993). The Echo Distributed File System. Technical Report 111, Digital Equipment Corp. Systems Research Center.

Boden, N., Cohen, D., Felderman, R., Kulawik, A., Seitz, C., Seizovic, J., and Su, W. (1995). Myrinet – A Gigabit-per-Second Local-Area Network. *IEEE Micro*, pages 29–36.

Bondurant, D. (1992). Enhanced Dynamic RAM. *IEEE Spectrum*, page 49.

Bursky, D. (1992). Memory-CPU Interface Speeds Up Data Transfers. *Electronic Design*, pages 137–142.

Cypher, R., Ho, A., Konstantinidou, S., and Messina, P. (1993). Architectural Requirements of Parallel Scientific Applications with Explicit Communication. In *Proceedings of the Twentieth International Symposium on Computer Architecture*, pages 2–13.

Dahlin, M., Mather, C., Wang, R., Anderson, T., and Patterson, D. (1994a). A Quantitative Analysis of Cache Policies for Scalable Network File Systems. In *Proceedings of the SIGMETRICS Conference on Measurement and Modeling of Computer Systems*, pages 150–160.

Dahlin, M., Wang, R., Anderson, T., and Patterson, D. (1994b). Cooperative Caching: Using Remote Client Memory to Improve File System Performance. In *Proceedings of the First Symposium on Operating Systems Design and Implementation*, pages 267–280.

Drapeau, A. (1993). *Striped Tertiary Storage Systems: Performance and reliability*. PhD thesis, University of California at Berkeley.

Gibson, G. A. (1992). *Redundant Disk Arrays: Reliable, Parallel Secondary Storage*. ACM Distinguished Dissertations. MIT Press, Cambridge, Massachusettes.

Gwennap, L. (1995). Processor Performance Climbs Steadily. *Microprocessor Report*.

Hart, C. (1992). Dynamic RAM as Secondary Cache. *IEEE Spectrum*, page 46.

Hartman, J. and Ousterhout, J. (1995). The Zebra Striped Network File System. *ACM Transactions on Computer Systems*.

Hennessy, J. and Patterson, D. (1996). *Computer Architecture A Quantitative Approach*. Morgan Kaufmann Publishers, Inc., 2nd edition.

Howard, J., Kazar, M., Menees, S., Nichols, D., Satyanarayanan, M., Sidebotham, R., and West, M. (1988). Scale and Performance in a Distributed File System. *ACM Transactions on Computer Systems*, 6(1):51–81.

Jones, F. (1992). A New Era of Fast Dynamic RAMs. *IEEE Spectrum*, pages 43–48.

Kazar, M. (1989). Ubik: Replicated Servers Made Easy. In *Proceedings of the Second Workshop on Workstation Operating Systems*, pages 60–67.

Keeton, K., Anderson, T., and Patterson, D. (1995). LogP Quantified: The Case for Low-Overhead Local Area Networks. In *Proceedings of the 1995 Hot Interconnects III Conference*.

Kistler, J. and Satyanarayanan, M. (1992). Disconnected Operation in the Coda File System. *ACM Transactions on Computer Systems*, 10(1):3–25.

Le, M., Burghardt, F., Seshan, S., and Rabaey, J. (1995). InfoNet: The Networking Infrastructure of InfoPad. In *Proceedings of COMPCON 95*, pages 163–168.

Liskov, B., Ghemawat, S., Gruber, R., Johnson, P., Shrira, L., and Williams, M. (1991). Replication in the Harp File System. In *Proceedings of the ACM Thirteenth Symposium on Operating Systems Principles*, pages 226–238.

Martin, R. (1994). HPAM: An Active Message Layer for a Network of HP Workstations. In *Proceedings of the 1994 Hot Interconnects II Conference*.

Martin, R. (1995). Personal Communication.

Mummert, L., Ebling, M., and Satyanarayanan, M. (1995). Exploiting Weak Connectivity for Mobile File Access. In *Proceedings of the ACM Fifteenth Symposium on Operating Systems Principles*, pages 143–155.

Ousterhout, J. (1990). Why Aren't Operating Systems Getting Faster As Fast As Hardware? In *Proceedings of the Summer 1990 USENIX Conference*.

Patterson, R., Gibson, G., Ginting, E., Stodolsky, D., and Zelenka, J. (1995). Informed Prefetching and Caching. In *Proceedings of the ACM Fifteenth Symposium on Operating Systems Principles*, pages 79–95.

Popek, G., Guy, R., Page, T., and Heidemann, J. (1990). Replication in the Ficus Distributed File System. In *Proceedings of the Workshop on the Management of Replicated Data*, pages 5–10.

Prince, B., Norwood, R., Hartigan, J., and Vogley, W. (1992). Synchronous Dynamic RAM. *IEEE Spectrum*, pages 44–46.

Rosenblum, M. and Ousterhout, J. (1992). The Design and Implementation of a Log-Structured File System. *ACM Transactions on Computer Systems*, 10(1):26–52.

Ruemmler, C. and Wilkes, J. (1993). UNIX Disk Access Patterns. In *Proceedings of the Winter 1993 USENIX Conference*, pages 405–420.

Sandhu, H. and Zhou, S. (1992). Cluster-Based File Replication in Large-Scale Distributed Systems. In *Proceedings of the SIGMETRICS Conference on Measurement and Modeling of Computer Systems*, pages 91–102.

Schroeder, M., Birrell, A., Burrows, M., Murray, H., Needham, R., Rodeheffer, T., Satterthwaite, E., and Thacker, C. (1991). Autonet: A High-Speed, Self-Configuring Local Area Network Using Point-to-Point Links. *IEEE Journal on Selected Areas in Communication*, 9(8):1318–1335.

Schroeder, M. and Burrows, M. (1990). Performance of Firefly RPC. *ACM Transactions on Computer Systems*, 8(1):1–17.

Seagate (1994). *ST-11200N SCSI-2 Fast (Barracuda 4) Specification*. Seagate Technology, Inc.

Seltzer, M., Bostic, K., McKusick, M., and Staelin, C. (1993). An Implementation of a Log-Structured File System for UNIX. In *Proceedings of the Winter 1993 USENIX Conference*, pages 307–326.

Thekkath, C. and Levy, H. (1993). Limits to Low-Latency Communication on High-Speed Networks. *ACM Transactions on Computer Systems*, 11(2):179–203.

von Eicken, T., Basu, A., and Buch, V. (1995). Low-Latency Communication Over ATM Networks Using Active Messages. *IEEE Micro*, pages 46–53.

Walker, B., Popek, G., English, R., Kline, C., and Thiel, G. (1983). The LOCUS distributed operating system. In *Proceedings of the ACM Ninth Symposium on Operating Systems Principles*, pages 49–69.